IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Approach Based On Machine Learning Is Utilized In The Developed Algorithm For Disease Prediction Model

[1]Prem Kumar Rathaor, [2]Manish Kumar Soni,

[1]M. Tech (Computer Science), Bansal Institute Engineering and Technology, Lucknow

[2]Assistant Professor, Department of Computer Science, Bansal Institute Engineering and Technology, Lucknow

Abstract: The method that is used to forecast illnesses based on the symptoms that are provided by patients or any user is referred to as Disease Prediction Using Machine Learning (or simply Disease Prediction). The user supplies the symptoms they are experiencing as the system's input, and the algorithm calculates the disease's likelihood based on those symptoms. The purpose of this article is to develop a dependable machine-learning model that is capable of accurately predicting the sickness that a human is suffering from based on the symptoms that the human is experiencing. Let us investigate the many ways in which we might tackle this topic via machine learning: Many facets of contemporary life have been profoundly influenced by machine learning (ML). It may be used for data mining to extract useful information from enormous datasets, for example, or for image analysis to tell one kind of picture from another. Discovering patterns, building models, and making predictions from training data are all within reach thanks to ML.

***Index Terms* - Machine Learning, Proposed Algorithms, Disease Prediction Model**

1.1 Introduction

Methods from the field of machine learning (ML) have been used in the process of analyzing big population datasets with the purpose of forecasting a variety of illnesses and determinants of risk. In the field of medical data analysis, logistic regression, abbreviated as LR, is one of the multivariate linear models that is used the most often. Making an accurate diagnosis of the patient's health is perhaps the most difficult component of working in the medical sector, [1]. The condition might be difficult to diagnose since doctors have to consider a considerable amount of clinical and pathological evidence before making a judgment. This intricacy has led to a surge in interest among healthcare providers and academics in the hunt for more accurate ways to forecast the onset of disease. Several well-known Data mining techniques have been developed and exploited in numerous real-world application areas (such as Industry, Healthcare, and Bio science) and are now being used in machine learning environments to extract useful pieces of information from the specified data in healthcare communities, biomedical fields, etc [2]. Reliable analysis of medical databases helps in early disease prediction, patient treatment, and community services. Disease prediction is only one area where machine learning methods have been effectively used.

1.2 Proposed Algorithms

In this section, a novel algorithm is proposed for breast tumor classification using a set of prediction steps. For the purpose of assessment, the methodology of publicly accessible machine learning is used, and current methods that are considered state-of-the-art are compared to it [3,4]. The input dataset is given to the various categorization learning models, as illustrated so that these models can recognize and discriminate between malignant and benign tumors by following the appropriate procedures. The performance of the kNN, SVM, and Tree classification methods, which are considered to be state-of-the-art machine learning algorithms, was evaluated and compared for the purpose of tumor classification.. If researchers in the domains of strategy and medical science are interested in understanding the concepts underlying machine learning for strategy research, as well as the newly developed applications and rising trends in the field, then they should read this article. It will provide them with information on all of the aforementioned topics and more. The field of machine learning has been witness to a significant convergence of ideas in recent years, which has resulted in the creation of a powerful new framework for the design of applications that are employed in the real world [5,6]. This chapter's goal is to call attention to the evolution of this new point of view and to highlight the numerous practical benefits it has in contrast to the ways that came before it. These goals will be accomplished by highlighting the many advantages this new point of view offers [7]. On the other hand, the objective of this chapter is not to offer complete coverage, and there is no attempt made to provide accurate historical credit for any of the many major contributions that have been made.

1.3 Algorithms

(SVM, kNN and Tree)

In: Inpu {Patient data};

Out: Criteria {Data Criteria at different level};

Step1: Data Preprocessing and Data Cleaning of disease specific data

Step2: For each variable Inpu data,

do Feature Selection of variable

Step 3: For each instance step 2 target columns and classify data using SVM, kNN and tree algorithm

Step 4: Determine the accuracy level MSE RMSE

Find confusion metrics

Step 5: Find Rank of instances

Step 6: Predict Disease Class as

Class=1 Positive=Normal State

Else if (class =0)

Negative= Variable State end for

Dataset value will be retrieved

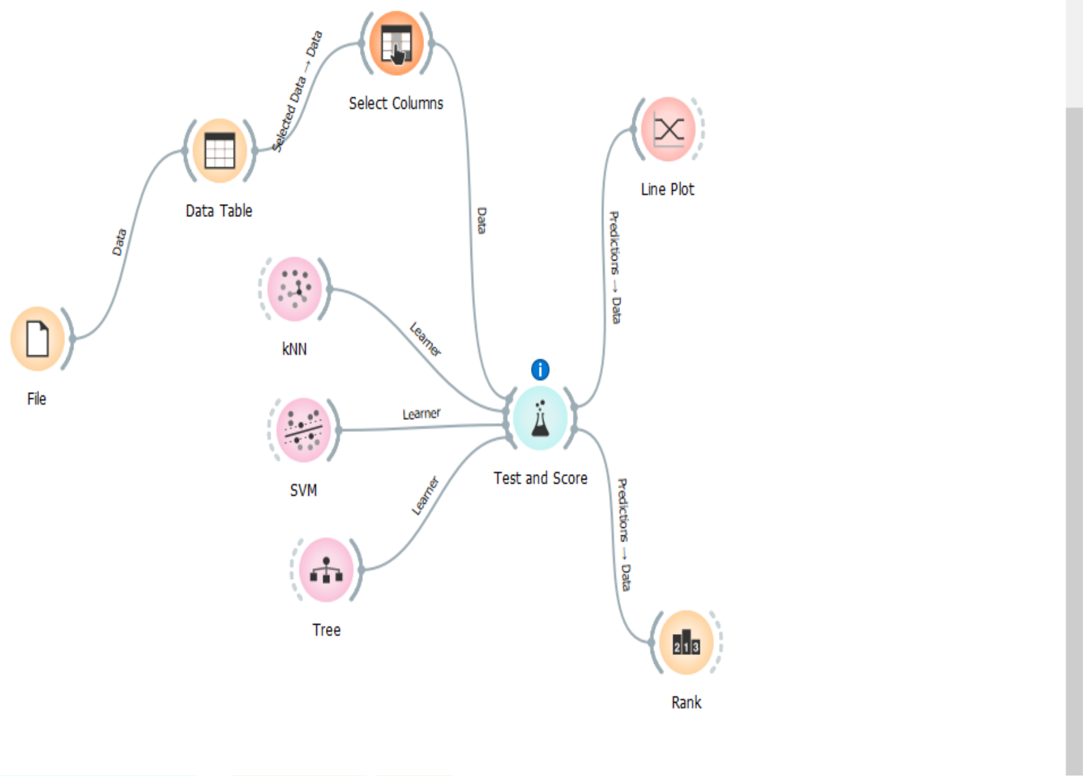


Fig 1.1 Data Flow

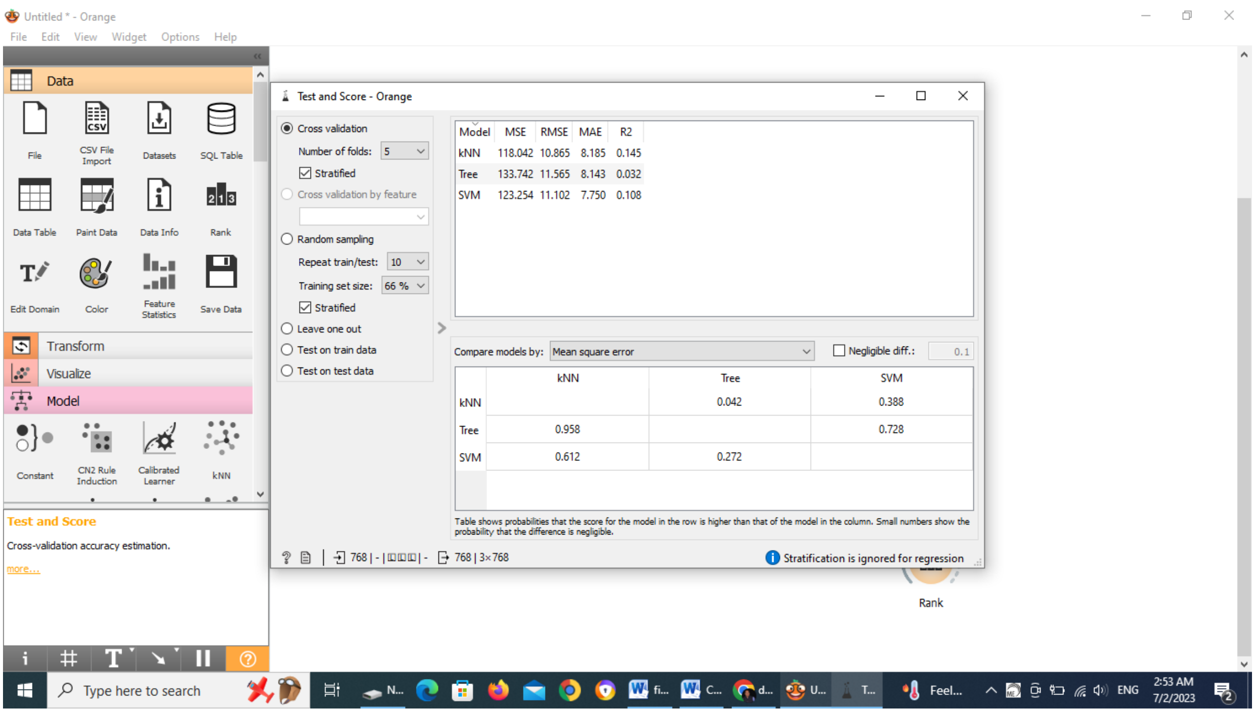


Fig 1.2: Model Comparison by MSE

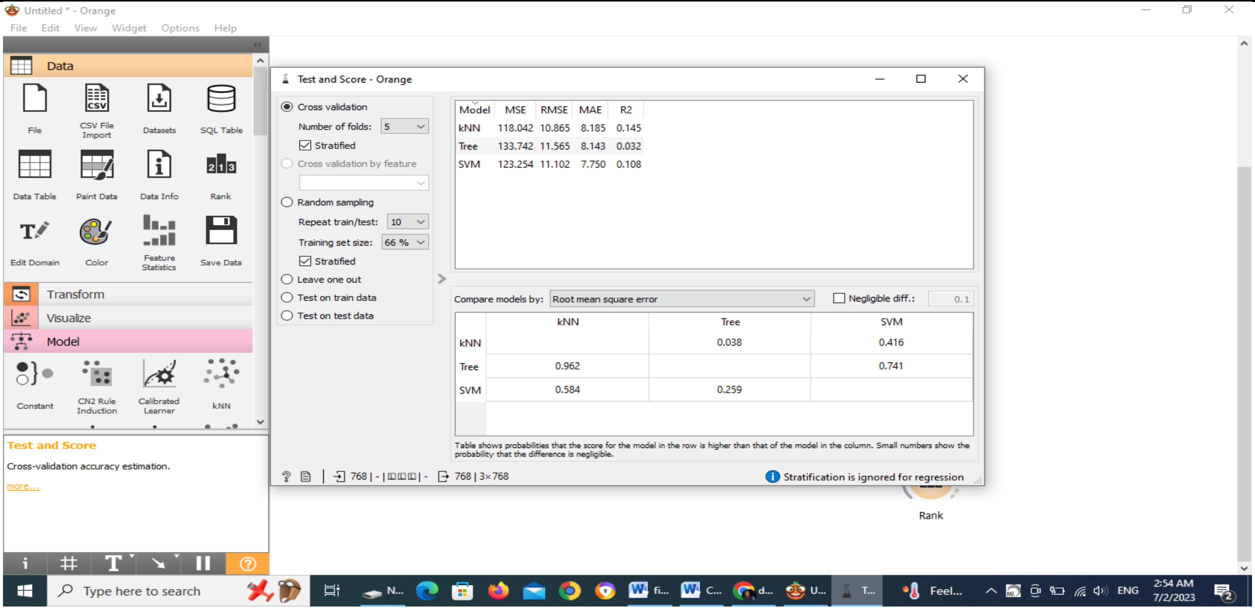


Fig 1.3: Model Comparison by RMSE

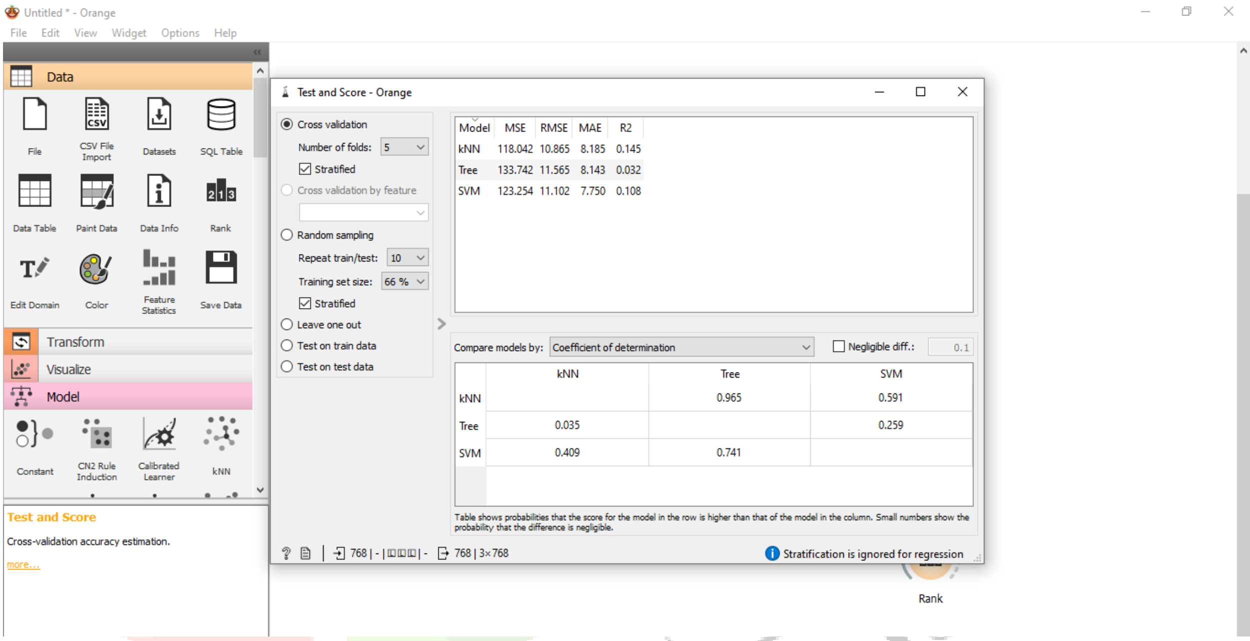
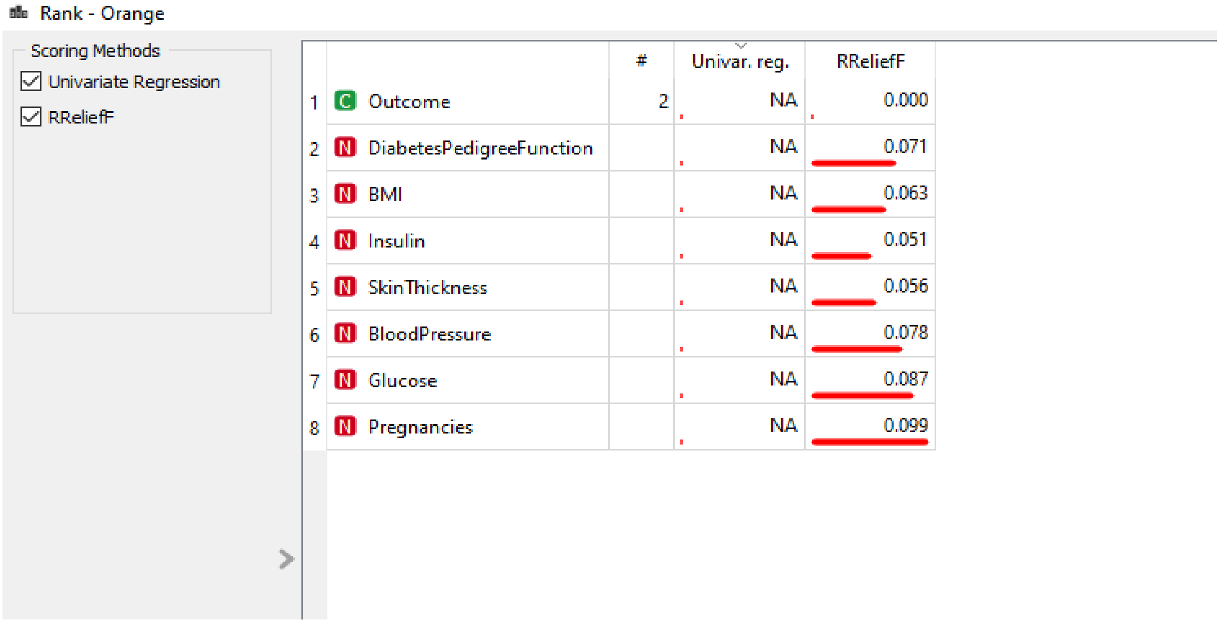


Fig 1.4: Model Comparison by Cofficient

The results of the machine learning algorithms performance' evaluation **(fig 1.1, 1.2 and 1.3)** using orange as the Machine Learning tool are depicted in **fig 1.2**. The study reveals that out of 3 machine learning algorithms namely **kNN, SVM and Tree** considered, **SVM (fig 1.3)** has the highest accuracy (0.998) followed by Sequential Minimal Optimization (SMO) and then Tree with 0.996 as shown in **fig 1.4**.

There is a close relationship between Accuracy and Precision. Beyond Accuracy, there are other metrics such as **AUC, CA, F1 and Precision and Recall**, Statistics that have a significant bearing on the decision of which method to use in model construction. This research highlights the significance of different criteria for performance assessment in connection to numerous Machine Learning approaches in addition to Accuracy in order to arrive at credible predictive analytics [8]. This is of the utmost importance since, as previously said, over-fitting may be deceiving. In addition to this, the heterogeneity of the dataset properties is another component that should be taken into consideration when looking at other measures.



Rank - Orange

Scoring Methods

- [x] Univariate Regression
- [x] RReliefF

		#	Univar. reg.	RReliefF
1	C Outcome	2	NA	0.000
2	N DiabetesPedigreeFunction		NA	0.071
3	N BMI		NA	0.063
4	N Insulin		NA	0.051
5	N SkinThickness		NA	0.056
6	N BloodPressure		NA	0.078
7	N Glucose		NA	0.087
8	N Pregnancies		NA	0.099

Fig 1.5 Rank of attributes

In fig 1.5, the things that we have looked at up to this point try to get rid of irrelevant qualities as well as duplicated ones. A more straightforward approach would be to rate the usefulness of every single characteristic, then choose the most important handful of those ratings to utilize for categorization and get rid of the rest. This method is very quick due to the fact that it does not require any searching at all, however it can only remove irrelevant properties and not redundant ones. And the outcomes, Glucose are quite sensitive to the total amount of characteristics that are kept.

1.4 Conclusion

This article made an effort to get an understanding of the many methods and strategies that are available for predicting the risk of diseases in humans by anticipating the risk variables. Considering the patient's health and reaction to treatment makes breast cancer diagnosis difficult. Machine learning improves breast cancer diagnosis. Despite technological breakthroughs, breast cancer diagnosis and monitoring remain difficult. Combining biological, sociological, and demographic data improves prediction models. SVM, kNN, and Tree effectiveness comparisons imply an ensemble method. Tree, kNN, and SVM machine learning models are used to fix the AHP pairwise comparison matrix discrepancy in this study. Training, validation, and testing simulations compare both techniques. SVM performs similarly to Tree in CR reduction, but it predicts unknown inputs more accurately and has a quicker convergence time than Tree.

References

1. Sidey-Gibbons, A. M. Jenni, and C. J. Sidey-Gibbons, “Machine learning in medicine: a practical introduction,” BMC Medical Research Methodology, vol. 19, 2019.
2. P. Mishra, V. Varadharajan, U. Tupakula, and E. S. Pilli, “A detailed investigation and analysis of using machine learning techniques for intrusion detection,” IEEE Communications Surveys & Tutorials, vol. 21, no. 1, pp. 686–728, 2019.
3. Mohan S., Thirumalai C., Srivastava G. Effective heart disease prediction using hybrid machine learning techniques. IEEE doi: 10.1109/ACCESS.2019.2923707.

4. Ahsan M.M., Mahmud M., Saha P.K., Gupta K.D., Siddique Z. (2019), Effect of data scaling methods on machine Learning algorithms and model performance. Technologies. 2021;9:52. doi: 10.3390/technologies9030052. 51.
5. Aljaaf A.J., Al-Jumeily D., Haglan H.M., Alloghani M., Baker T., Hussain A.J., Mustafina J. Early prediction of chronic kidney disease using machine learning supported by predictive analytics; Proceedings of the 2018 IEEE Congress on Evolutionary Computation (CEC); Rio de Janeiro, Brazil. 8–13 July 2018; pp. 1–9.
6. HeonGyu Lee, Ki Yong Noh, KeunHoRyu, "Mining Biosignal Data: Coronary Artery Disease Diagnosis using Linear and Nonlinear Features of HRV" LNAI 4819: Emerging Technologies in Knowledge Discovery and Data Mining, pp. 56-66, May 2007.
7. Niti Guru, Anil Dahiya, NavinRajpal, "Decision Support System for Heart Disease Diagnosis Using Neural Network", Delhi Business Review, Vol. 8, No. 1 (January - June 2007).
8. Latha Parthiban and R.Subramanian, "Intelligent Heart Disease Prediction System using CANFIS and Genetic Algorithm", International Journal of Biological, Biomedical and Medical Sciences, Vol. 3, No. 3, 2008.